



ParaWise/CAPO Parallelization Environment

Henry Jin

hjin@nas.nasa.gov

NASA Advanced Supercomputing (NAS) Division

NASA Ames Research Center

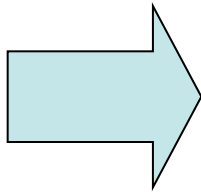
June 13, 2005

Key Ideas

- **Interactive** environment for semi-automatic parallelization of Fortran application codes
- Generated codes in **recognizable form** by user

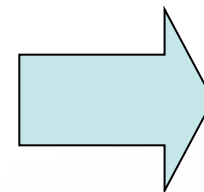
INPUT

Fortran
code



ParaWise/CAPO

Transformation
Parallel code



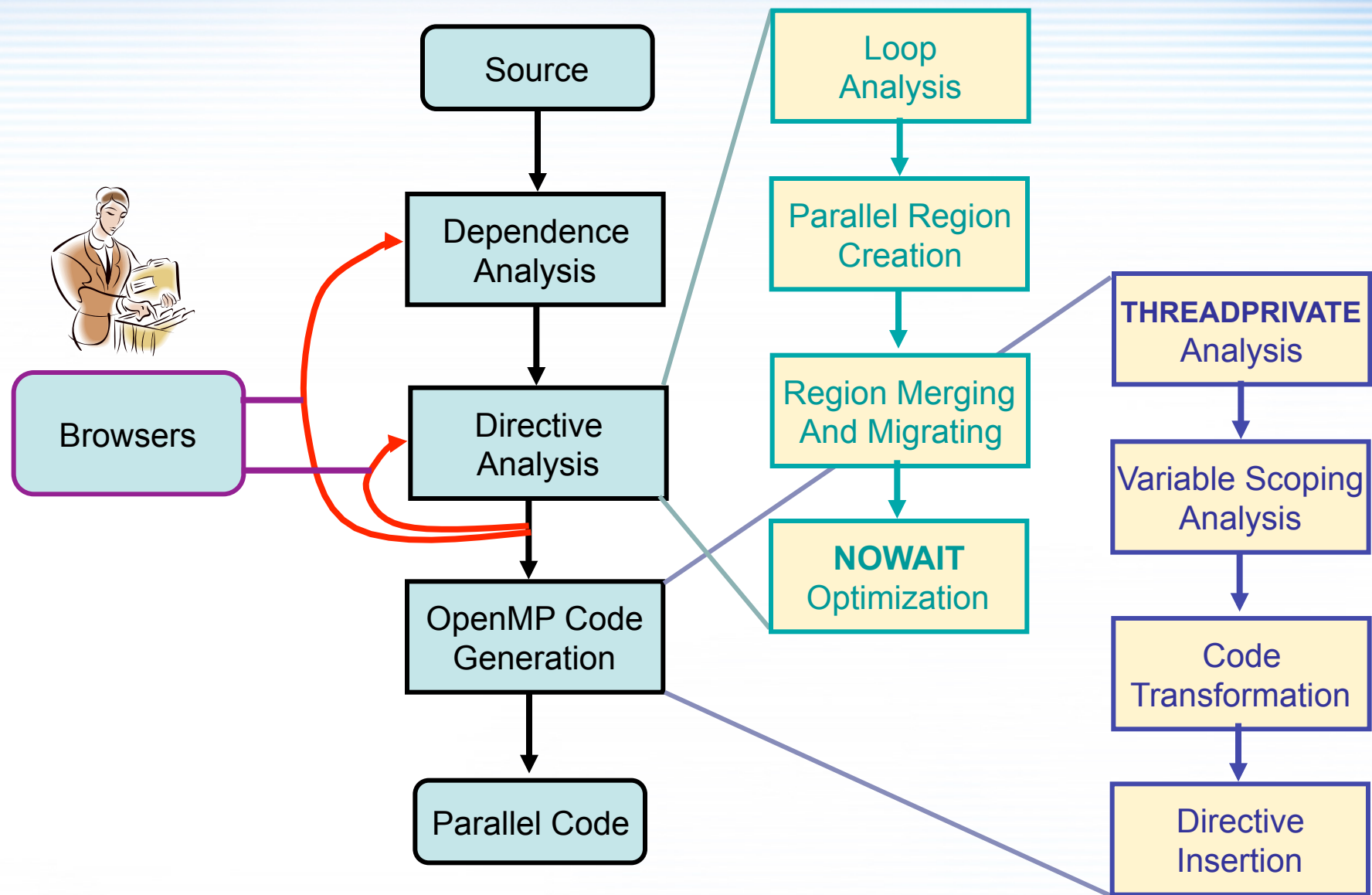
OUTPUT

Fortran +
OpenMP
directives

ParaWise and CAPO

- ParaWise
 - Semi-automatic, developed by Parallel Software Products
 - Accurate symbolic, value based, interprocedural data dependence analysis
 - Domain decomposition for generating message-passing codes
 - A set of browsers for user to interact with the parallelization process
- CAPO
 - A module for generating OpenMP parallel codes, developed at NASA Ames
 - Exploits loop-level parallelism
 - Directives browsers to guide the parallelization process
 - Currently integrated with ParaWise

Interactive Parallelization Process



Generation of OpenMP Code

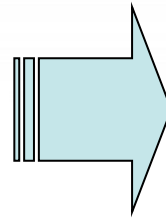
- Identify parallel loops, including loops for setting up possible pipeline
- Construct parallel regions from parallel loops
- Merge consecutive parallel regions and migrate parallel regions as high as possible in the call path
- Perform **NOWAIT** optimization for consecutive parallel loops inside a parallel region
- Automatically identify and define variable scopes, such as **SHARED**, **PRIVATE** and **REDUCTION**
- Detect and produce **THREADPRIVATE** directives for common blocks

Code Generation Process

serial code

```
do K=
  ...
end do
call subwork
...

subroutine subwork
do J=
  ...
end do
do J=
  ...
end do
return
end
```



```
!$OMP PARALLEL DO

do K=
  ...
end do
!$OMP END PARALLEL DO
call subwork

...

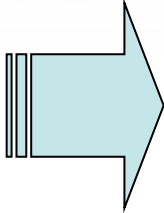
subroutine subwork

!$OMP PARALLEL DO
do J=
  ...
end do
!$OMP END PARALLEL DO
!$OMP PARALLEL DO
do J=
  ...
end do
!$OMP END PARALLEL DO

return
end
```

*identify parallel loops
create parallel regions*

Code Generation Process (cont.)



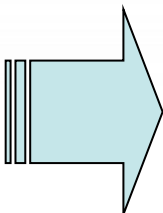
```
!$OMP PARALLEL DO

do K=
...
end do
!$OMP END PARALLEL DO
call subwork

...

subroutine subwork
!$OMP PARALLEL
!$OMP DO
do J=
...
end do
!$OMP END DO
!$OMP DO
do J=
...
end do
!$OMP END DO
!$OMP END PARALLEL
return
end
```

merge parallel regions



```
!$OMP PARALLEL
!$OMP DO
do K=
...
end do
!$OMP END DO
call subwork
!$OMP END PARALLEL
...

subroutine subwork

!$OMP DO
do J=
...
end do
!$OMP END DO NOWAIT
!$OMP DO
do J=
...
end do
!$OMP END DO NOWAIT

return
end
```

*migrate parallel regions
generate NOWAIT*

Automatic Code Transformation

- Privatization of common block variables
 - if cannot be handled with **THREADPRIVATE**
- Routine duplication
 - to resolve conflicts of usage
- Reduction on an array variable
 - update local variable in parallel, then the shared array variable in a critical region
- F90 array syntax to loop nest
 - so that **OMP DO** can be applied
- Loop interchange
 - for better cache utilization

Routine Duplication

- Call inside a parallel region, but not inside a parallel DO

inside parallel region

outside parallel region

```
call sub
do K=
...
end do
...
call sub

subroutine sub
do J=
...
end do
```

```
!$OMP PARALLEL
  call sub
!$OMP DO
  do K=
    ...
  end do
!$OMP END PARALLEL
...
call sub

subroutine sub
!$OMP PARALLEL DO
  do J=
    ...
  end do
```

```
!$OMP PARALLEL
  call cap_sub
!$OMP DO
  do K=
    ...
  end do
!$OMP END PARALLEL
...
call sub

subroutine sub
!$OMP PARALLEL DO
  do J=
    ...
  end do

subroutine cap_sub
!$OMP DO
  do J=
    ...
  end do
```

Identifying Parallel Loops - The Key Issue

- *Code developers want to*
 - find all the loops that can be parallelized
 - find all those that look ‘serial’
 - find which of the ‘serial’ don’t affect parallel performance and which are critical
 - fix the code so that the critical ‘serial’ loops can be parallelized
- *CAPO enables this function by*
 - categorizing different loop types
 - solving through user interaction
 - generating parallel code with directives automatically

Directives Browser Window

CAPO: Directives Browser

Scope: All Routines

Loop Filter: **Covered Serial** (highlighted with a red circle and arrow labeled "loop type")

Sub Filter: All

36 Routines: NBSH20, NBSLOX, NEWVIS, PROPTY, RVA4, SOLVEP, SOLVEQ, SOLVES, SOLVET, SOLVEU, SOURCE

8 Covered serial loops (True dependence, containing parallel loop)

4/2/117: DO 991 II=1, IZON, 1
 6/3/138: DO 991 III=1, IDIM, 1
7/4/158: DO 331 K=KS, KT, 1
 8/5/159: DO 331 J=2, JT-1, 1
 34/2/585: DO 1991 II=1, IZON, 1
 38/3/600: DO 1991 III=1, IDIM, 1
 39/4/620: DO 1331 K=KS, KT, 1

More Filter... More Browsers: Region... Array Syntax... Routine Dup...

Current Routine: SOLVEP

```

151 :      endif
152 :      if (IDIM.eq.2) THEN
153 :          KS=1
154 :          KT=1
155 :          KX=0
156 :      endif
157 : C****(I-FLUX)*****
158 :      DO 331 K=KS,KT,1
159 :          DO 331 J=2,JT-1,1
160 : C-----START I-LINE
161 :          DO 301 i=1,IT,1
162 :              if (III.eq.1) THEN
163 :                  IQP(i)=L+i+(J-1)*
164 :              endif
165 :              if (III.eq.2) THEN
166 :                  IQP(i)=L+J+(i-1)*
167 :              endif
168 :              if (III.eq.3) THEN
169 :                  IQP(i)=L+J+(K-1)*
170 :              endif
171 :              ijk=IQP(i)
172 :              qqp(i,14)=du(ijk)*a
173 :          301 CONTINUE
174 : C-----INTERFACE VELOCITY
175 :          DO i=2,IT,1
176 :              ijk=IQP(i)
177 :              ip1=MAX0(0,MIN0(1,
178 :              im1=MAX0(0,MIN0(1,
179 :              FFG(i)=1.0*ip1*im1
    
```

CAPO: Why Directives ?

Loop: 7/4/158: DO 331 K=KS,KT,1 Type: Covered Serial

Reason: True dependence, containing parallel loops

True-dep. variables: FFA, FFB, FFD, FFE, SU

Anti-dep. variables: FFA, FFB, FFC, FFD, FFE, FFF, FFG, ffi

Output-dep. variables: AE, AW, FFA, FFB, FFC, FFD, FFE, FFF

In/out-dep. variables: >AE, >AW, <FFA, <FFB, >FFB, <FFC, >FFC

Update Directives... Reset DeLoop... Privatize... Remove... AE, AW, SU

IO/Exit statements:

Hints:

- Contains 15 parallel loops
- 5 variables with loop-carried true dependence (level=4)
- 12 variables with loop-carried anti dependence (level=4)
- 14 variables with loop-carried output dependence (level=4) and non-privatizable, due to usage from outside the loop

Contains 15 parallel loops:

```

SOLVEP:9/6/161: DO 301 i=1,IT,1
SOLVEP:10/6/175: DO i=2,IT,1
SOLVEP:11/6/188: DO i=1,IT,1
SOLVEP:12/6/226: DO i=2,IT,1
    
```

Inside parallel loops:

Loop Types Identified with Directives Browser

Totally Serial

Problem : Potentially severe

- Serial due to loop-carried true dependence present and/or,
- Serial due to loop-carried pseudo (memory re-use) dependence by a non-privatizable variable
- Not contained in, or containing ANY parallel loops - entirely serial
- Sequential execution can prevent effective parallel performance

Possible Solutions :

- True dependence may have been assumed, may be proven to no longer exist if user knowledge is added.
- Investigate loop-carried pseudo dependence - add user knowledge to prove non-existence.
- Investigate privatization preventing true dependences from/to outside of loop - add user knowledge to prove non-existence

Browser shows serializing dependences (textually and graphically)

Loop Types Identified with Directives Browser (cont.)

Covered Serial

Problem: May be important

- Also a serial loop, but contains or is contained in a parallel loop so some parallelism will be exploited.
- If contains parallel loops, parallel performance can be enhanced by parallelism at this higher level.

Possible solutions :

- Can be treated in a similar manner to the “serial” loop type described previously.

Browser shows serializing dependences and surrounding parallel loop(s) and/or contained parallel loops

Loop Types Identified with Directives Browser (cont.)

Chosen Parallel :

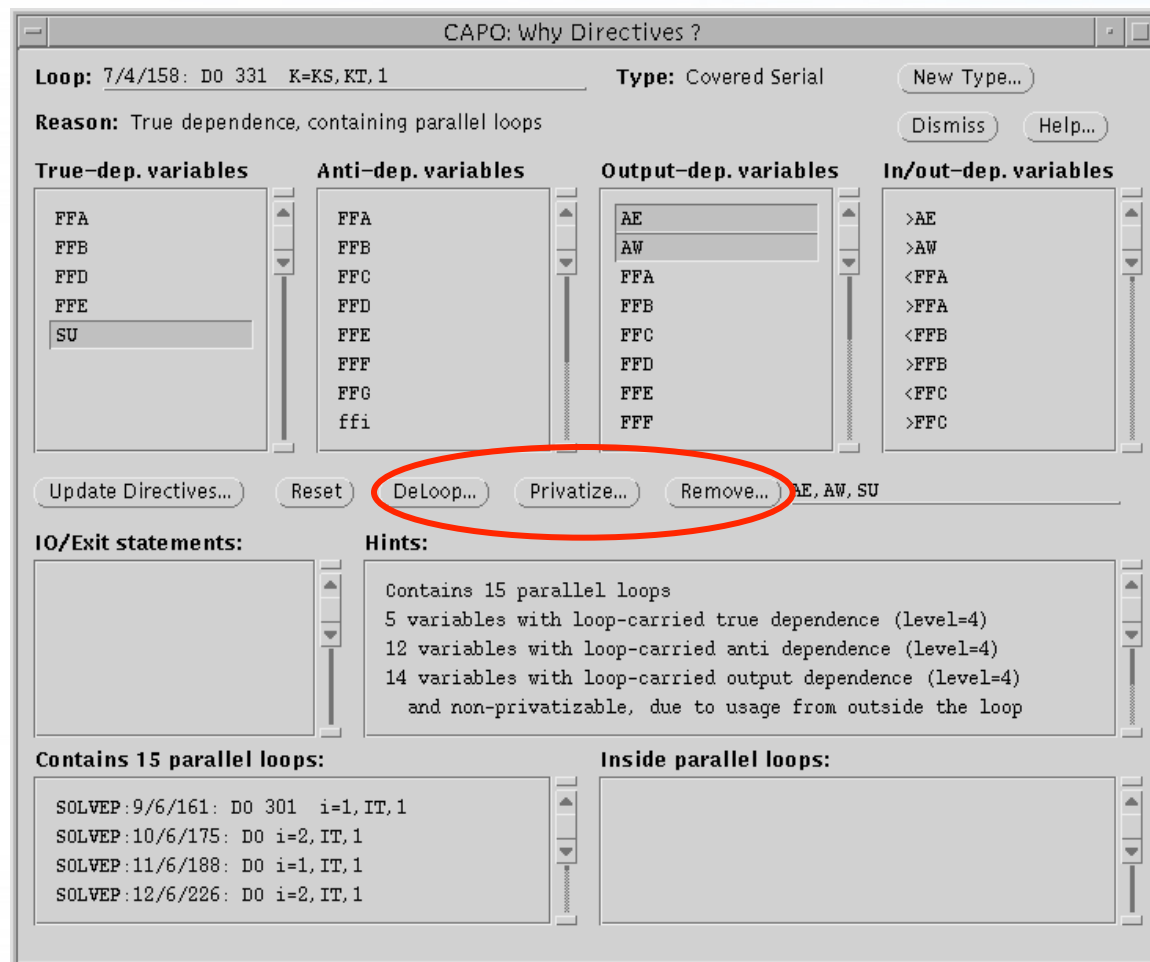
- Parallel loop that is not nested within other parallel loops
- Current Loop level at which parallel DO directive is inserted
- Includes loops identified with reduction operations
- Includes loops identified with software pipelines

Not Chosen:

- Parallel loop not chosen due to the selection of other parallel loops from the “Chosen Parallel” category above or due to I/O statements
- User may enforce parallelization if needed

The Why Directives Window

- Reason and hints for a selected loop
- List of variables and dependence types
- Tools for removing dependences



Investigate Why a Dependence Is Defined

Statements...

Dependence Filter...

Normalize Text

Properties...

Tools ▾

Graph ▾

History ▾

Why Dependence ?

ReDraw

Dismiss

Help...

Current Routine: TDMA

Current Graph: 7 statements, 13 dependencies displayed

1 : SUBROUTINE TDMA(Y, IN, IN1, J)

2 : C

3 : C SOLVES A TRIDIAGON

4 : C

5 : REAL Y(500,500),A(

6 : INTEGER IN, IN1, I1,

7 : COMMON /TDMARR/A,B

8 : B(2)=B(2)-A(2)*Y(1

9 : B(IN1)=B(IN1)-C(IN

10 : C

11 : C FORWARD ELIMINA

12 : C

13 : DO 10 I=3, IN1, 1

14 : I1=I-1

15 : M1=A(I)/D(I1)

16 : B(I)=B(I)-M1*B(I

17 : D(I)=D(I)-M1*C(I

18 : 10 CONTINUE

19 : C

20 : C BACK SUBSTITUTI

21 : C

22 : Y(IN1, J)=B(IN1)/D(

23 : DO 20 I=IN-2, 2, -1

24 : Y(I, J)=(B(I)-Y(I

25 : 20 CONTINUE

26 : RETURN

27 : END

ParaWise: Why Dependence ?

Type: TRUE

Level: Loop Independent

In Routine: TDMA

Caused By Array: B

Domination: PRE POST

Source Line: 8:B(2)=B(2)-A(2)*Y(1,J)

Sink Line: 9:b(in1)= B(IN1) -c(in1)*y(in,j)

During Iterations Of Loop:

Array Index: 1

Couldn't Disprove: 2 = IN1

Dependence Status: Can Be Removed If Any Question Below Is True

2 Questions:

1 : IN <= 2

1 : IN >= 4

Defining Statement Call Path:

CONDUC:13:CALL GEOMET(Z,R,RSTRT)

FAB:14:CALL CONDUC(RSTRT,TIME,TL,DT)

Defining Statements For Variables In Question:

IN:GEOMET:13:READ(10,*)IN,JN,IMON,JMON,HGT,WITH

Question Is Definitely True

Remove This Dependence

Dismiss

Help...

Recursive Sources...

Recursive Sinks...

1 : SUBROUTINE GEOMET(Z,R,RSTRT)

2 : C

3 : C THIS ROUTINE SETS UP THE GRID GEOMETRY SPECIFICATION AS (IN,JN)

4 : C

5 : REAL DZ,VAL,WITH,RIN,HGT,Z(500),R(500),FACK,FACY

6 : INTEGER GMOPT,IN,JN,LOC,IMON,JMON,RSTRT

7 : CHARACTER*18 LAB,XLAB0,XLAB1,YLAB0,YLAB1,XNAM0,XNAM1,YNAM0,YNAM1

8 : CHARACTER ANS,XLABMON,YLABMON

9 : PARAMETER (FIXVAL=1.0E+10,ADIABAT=1.0E-10)

10 : PARAMETER (XLAB0='X',XLAB1='Z',YLAB0='Y',YLAB1='R',XNAM0='HEIGHT',XNAM1

11 : = 'RADIAL THICKNESS',YNAM0='WIDTH',YNAM1='LENGTH OF CYLINDER')

12 : COMMON /GEOM/IN,JN,GMOPT,RIN,WITH,HGT,IMON,JMON,FACK,FACY

13 : READ(10,*)GMOPT,RIN

14 : READ(10,*)IN,JN,IMON,JMON,HGT,WITH

15 : READ(10,*)FACK,FACY

16 : IF (ABS(FACK-1.0).LE.ADIABAT) THEN

17 : DZ=WITH/(IN-2)

18 : ELSE

19 : F1=FACK*(IN-2)

20 : DZ=2.0*WITH*(1.0-FACK)/(1.0+FACK-F1+FACK)

21 : ENDIF

22 : C

Further Code Optimization

- Choose outer-most loops for better granularity
- Prune data dependences when
 - unknown information involved (e.g. input parameters)
 - code too complicated (e.g. FFT)
- Require user knowledge
- Use dialog boxes in the WhyDirectives window
 - remove **false** data dependences
 - thus parallelize a loop

Remove Data Dependencies

```
integer indexptr(maxcells)
read*,indexptr
do i=1,ncells
  S1    u(indexptr(i)) = . . .
  S2    . . . = u(indexptr(i)) + . . .
enddo
S3    print*, (u(j), j=1,ncells)
```

analysis

- i loop serial due to loop carried pseudo dependences of u , $S_1 \rightarrow S_1$ (**output**), $S_2 \rightarrow S_1$ (**anti**), Loop output $S_1 \rightarrow S_3$ also u is not PRIVATE

user inspection

- Examine Loop output dependence and determine it is correct therefore u cannot be PRIVATE

possible solution

- If contents of `indexptr` are all unique then we can safely remove the loop carried **anti** and **output** dependencies for the array u allowing u to stay SHARED and the loop to execute in parallel

Remove Data Dependences (cont.)

```
S1  read*, (work(k), k=1,10)
      do i=1,10
        do j=1,n
S2          work(j)=j
        enddo
S3    b(i)=b(i)+work(2)
      enddo
```

analysis

- i loop serial due to loop carried pseudo dependencies of `work`, $S_2 \rightarrow S_2$ (**output**), $S_3 \rightarrow S_2$ (**anti**)
- Loop input dependence of `work`, $S_1 \rightarrow S_3$ (**true**) exists so `work` is not PRIVATE

user inspection

- Examine in Why dependence window of dependence graph browser
- Determine that the pseudo dependencies are correct (`work` is re-used)
- Loop input dependence non-existent if $n \geq 2$

possible solution

- Delete loop input dependence or (preferably) add $n \geq 2$ to info + re-analyze. `work` is now PRIVATE and i loop can execute in parallel

Remove Data Dependences (cont.)

```
S1  read* , (work(k) , k=1 , 10) , (n(k) , k=1 , 10)
      do i=1 , 10
        do j=1 , n(i)
S2          work(j)=j
        enddo
S3    b(i)=b(i)+work(2)
      enddo
```

analysis

- Now n is an array – additional **true** dependence of work carried by i loop $S_2 \rightarrow S_3$
- i loop appears to be inherently serial

user inspection

- Examine **true** dependence first, others only important if it can be removed
- Loop carried **true** dependence non-existent if all $n(1:10) \geq 2$

possible solution

- Delete loop carried **true** dependence followed by loop input dependence (as before) or just add $n(1:10) \geq 2$ to info + re-analyze
- i loop is now parallel and `work` is `PRIVATE`

Remove Data Dependences (cont.)

```
do k=1,d(3)
  do j=1,d(2)
    do i=1,d(1)
      S1      y1(j,i)=. . .
    enddo
  enddo
  S2      call cfftz(y1, . . .)
  do j=1,d(2)
    do i=1,d(1)
      S3      . . . = y1(j,i)
    enddo
```

analysis

- k loop is apparently serial since **y1** is assigned in S_1 and S_2 and is used in S_2 and S_3 i.e. **true** dependence $S_2 \rightarrow S_2$

user inspection

- Examine **true** dependence first, others only important if it can be removed.
- Examine loop Input/Output dependence

possible solution

- If it is known that there are no assignments of **y1** before S_1 then we can safely remove the loop carried **true** dependences and Input/Output dependences for **y1** making it

Remove Data Dependences (cont.)

- [DeLoop]
 - make variables **shared** → delete **loop-carried** dependences
- [Privatize]
 - make variables **private** → delete loop-carried **True/Anti** dependences and **Input/Output** True dependences

CAPO: VarList DeLooping

Unselected Variables:		Selected Variables:	
FFA	: T A O -	SU	: T A O -
FFB	: T A O -	AE	: - - O -
FFD	: T A O -	AW	: - - O -
FFE	: T A O -		
FFC	: - A O -		
FFF	: - A O -		
FFG	: - A O -		
ffi	: - A O -		
ffj	: - A O -		

Remove I/O Dependence: ☒ Both ☐ Input ☐ Output

Click Apply to remove Loop-Carried Dependences for the selected variables !

Reset

Apply Dismiss Help...

CAPO: VarList Privatization

Unselected Variables:		Selected Variables:	
SU	: T A - <>	FFA	: T A - <>
IQP	: - A - <>	FFB	: T A - <>
qqp	: - A - <>	FFD	: T A - <>
AE	: - - - >	FFE	: T A - <>
AW	: - - - >	FFC	: - A - <>
		FFF	: - A - <>
		FFG	: - A - <>
		ffi	: - A - <>
		ffj	: - A - <>

Remove I/O Dependence: ☒ Both ☐ Input ☐ Output

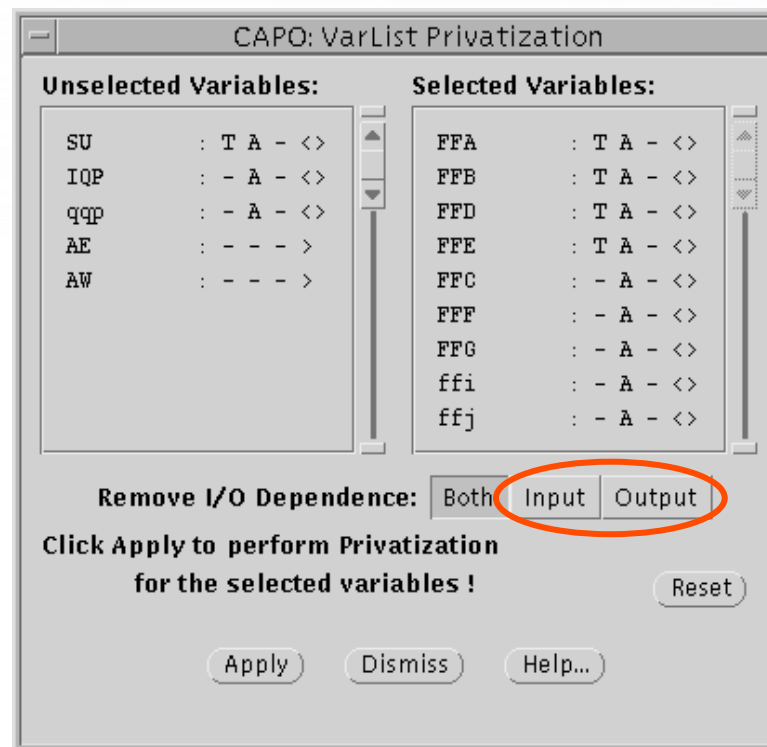
Click Apply to perform Privatization for the selected variables !

Reset

Apply Dismiss Help...

Remove Data Dependences (cont.)

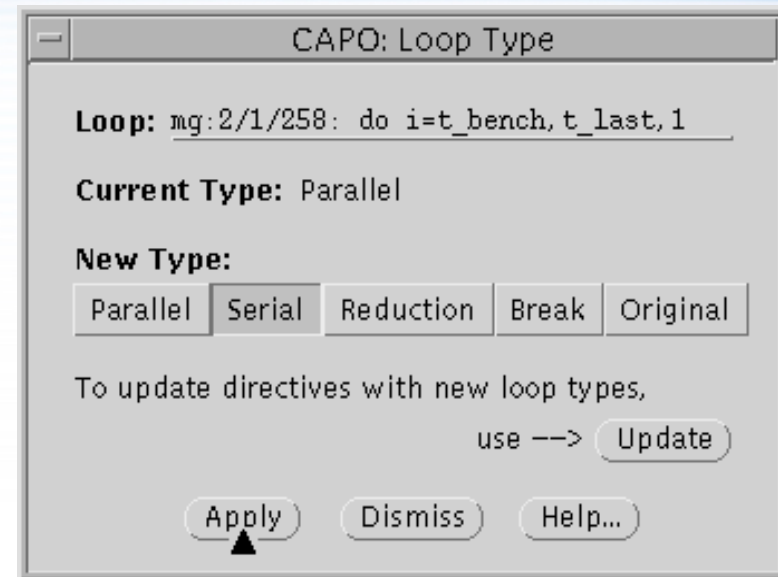
- [Privatize] continued
 - It is possible to make variables **firstprivate** or **lastprivate** → select “Remove **Output** (>) or **Input** (<) dependences”



Caution! User can improve performance **but also** can introduce mistakes

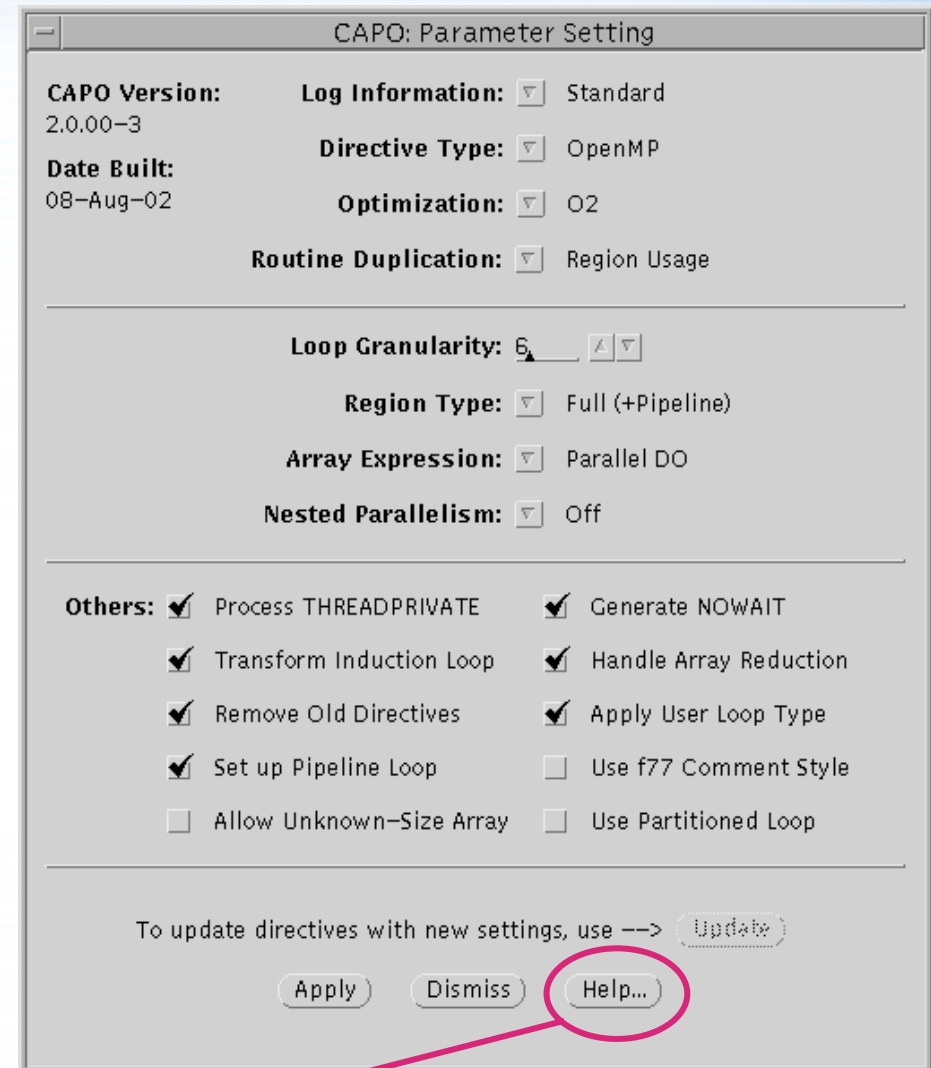
Further Optimization

- User enforced loop types
 - overwrite a default
 - for I/O loops
 - concerning granularity
 - use the Loop Type window
- The “userloop.par” file
 - User defined loop types are saved to this file, read back automatically from the file
 - A different filename may be specified via the environment variable **CAPO_USERLOOP**



Parameters to Control the CAPO Execution

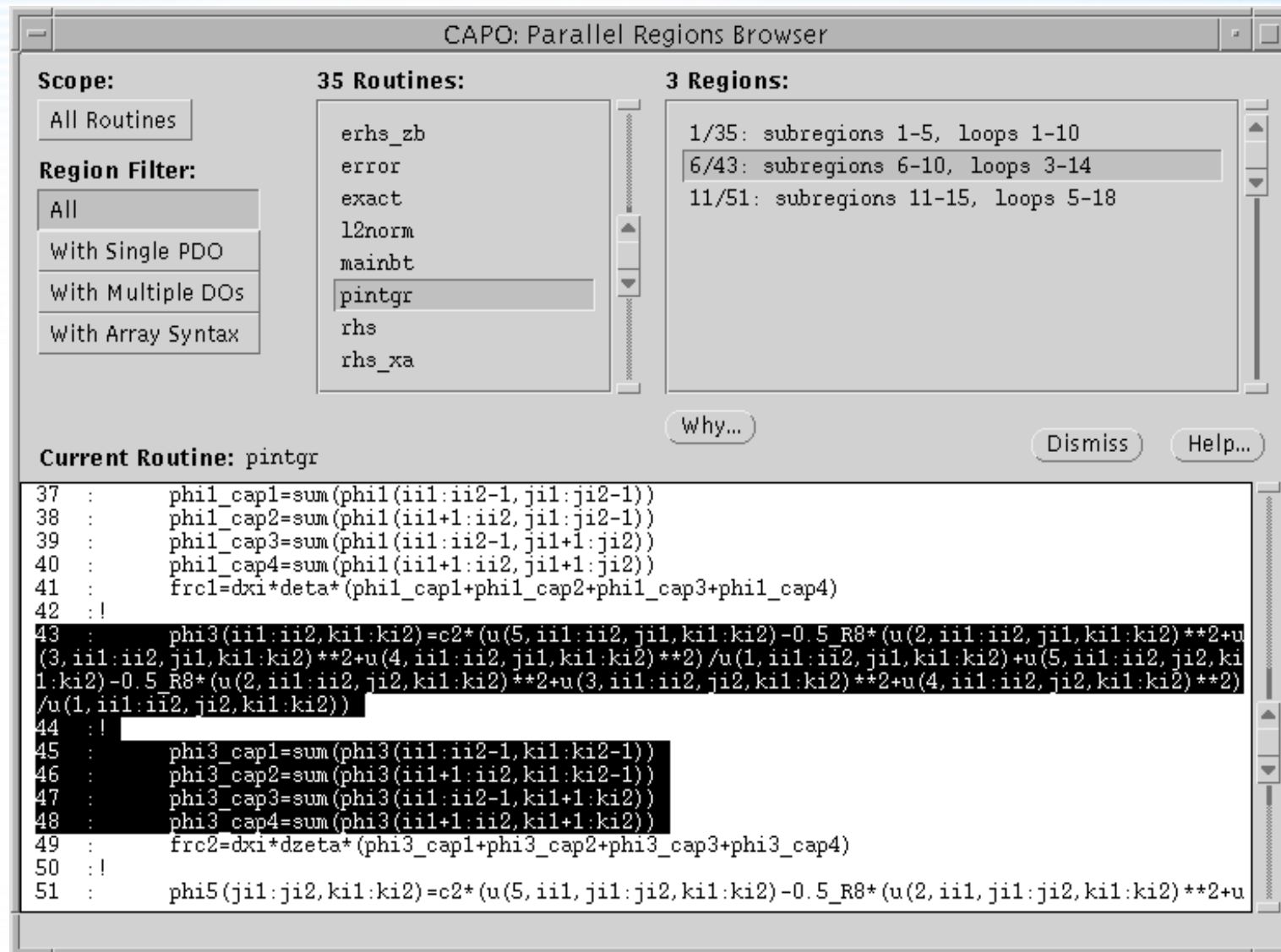
- Setting dialog box
 - set most parameters
- Environment variables
 - GUI correspondence
 - CAPO_LOGINFO
 - ...
 - no GUI correspondence
 - CAPO_USERLOOP
 - ...



if you are not sure

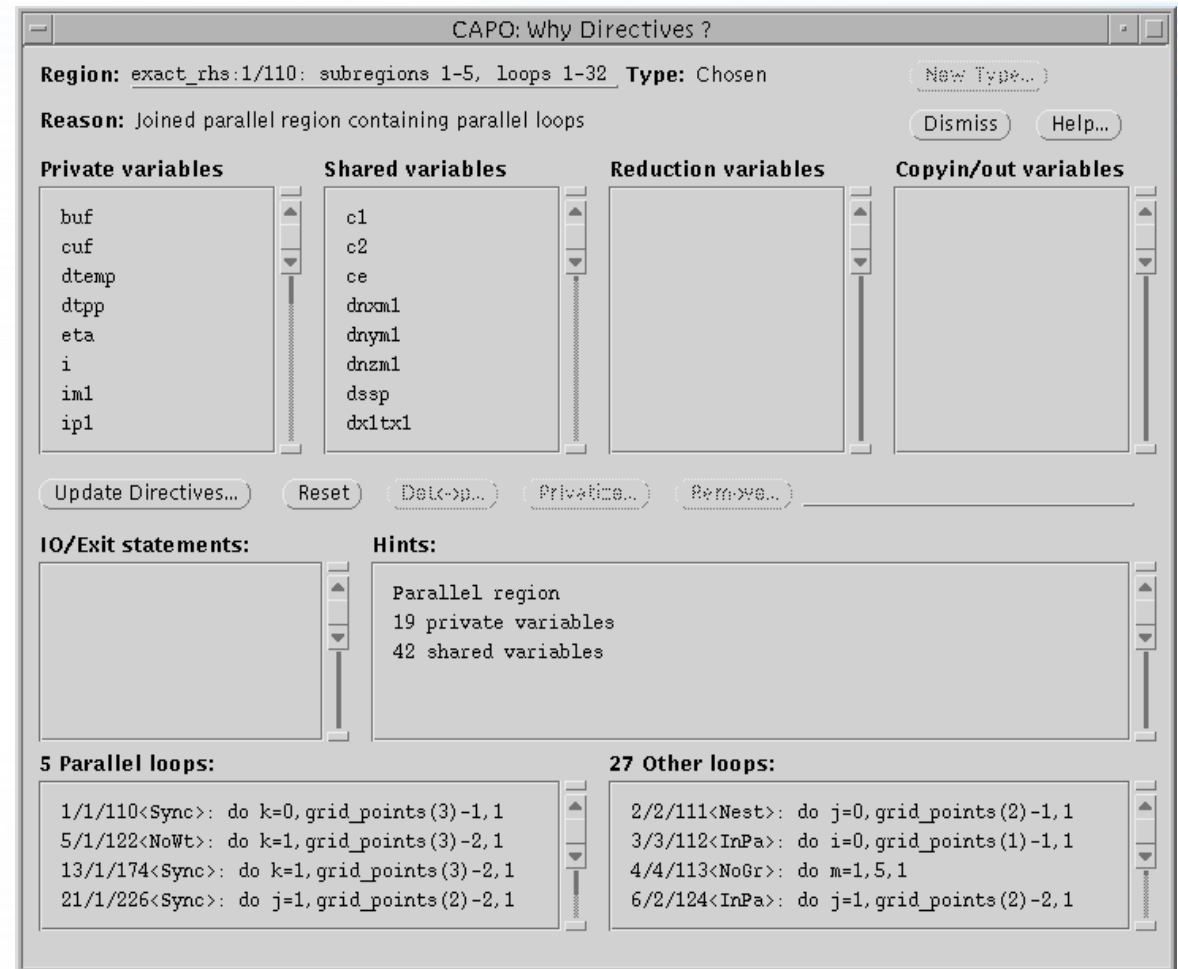
Browsing Parallel Regions

- With the *Parallel Regions* browser



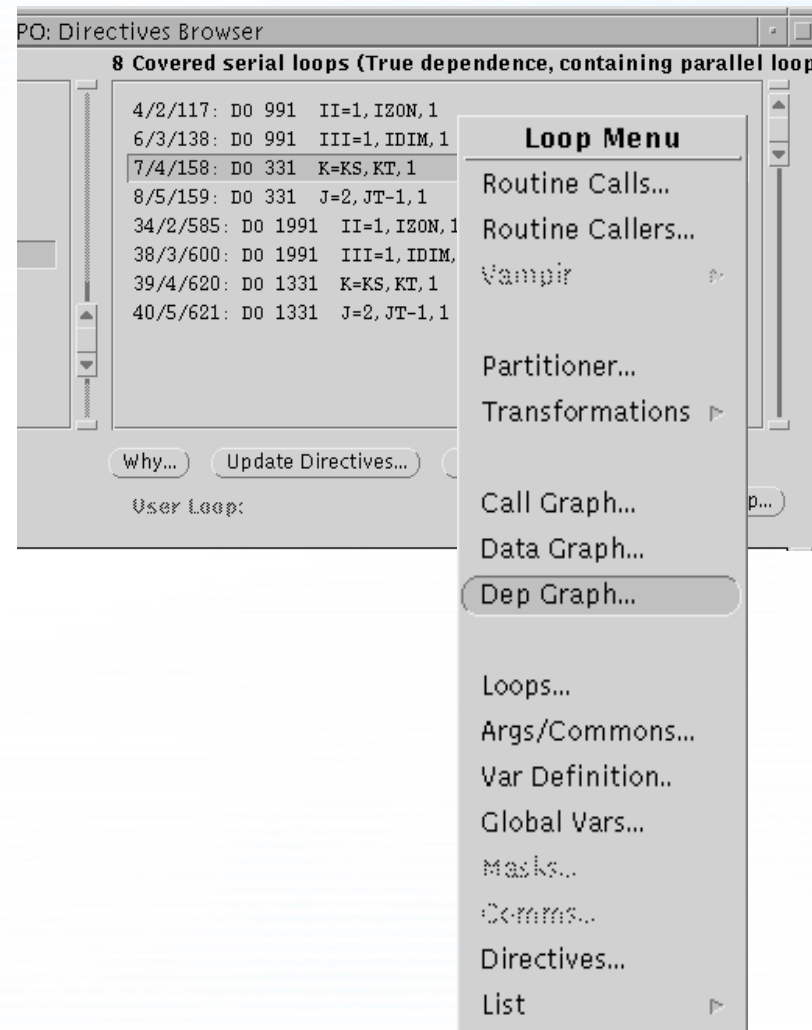
Browsing Parallel Regions (cont.)

- Connection to the WhyDirectives window
 - list of variables and their types
 - indication of the end-of-loop synchronization
- No direct modification to regions



Hotlinks

- Quick access to other functions
- Menus from pressing the *right* mouse button
 - linked with a loop
 - linked with a variable
 - linked with a routine
 - linked with a textline
- Example
 - bring up the DepGraph window for the selected loop



Command Interface for the Batch Mode

- Provide access to the functionality of GUI components without starting the GUI
- Commands usually recorded to a command file by
`capo -logfile capo_run.cmd`
- Played back [in a batch mode] with
`capo [-batch] capo_run.cmd`
- Commands in the command interface are given in the *CAPO User Manual A4*

Hybrid Parallelization

- Existing message passing codes
 - Use CAPO to insert OpenMP directives
- Two-step process
 - First: ParaWise to generate the message-passing code
 - Second: CAPO to insert OpenMP directives
- Issues
 - No communication routines allowed inside a parallel region
 - The partitioned dimension is not used for OMP loop level parallelization, but it is possible to enforce the choice
 - In the Setting Box, check “Use Partitioned Loop”
- See an example in the CAPO tutorial notes

Fortran 90/95 Codes

- In the beta stage
- Main feature – handling array syntax, **FORALL** loop, **WHERE** construct
 - convert to a regular DO loop
 - use “**OMP WORKSHARE**” (not yet supported)
 - do nothing, let a compiler work it out

```
flux(2,2:nx-1,2:ny-1,2:nz)=tz3*(du2(2:nx-1,2:ny-1,2:nz) &  
&                                -du2(2:nx-1,2:ny-1,1:nz-1))
```

converted to

```
do ARRAY_VAR3=2,nz  
  flux(2,1:nx,2:ny-1,ARRAY_VAR3)=tz3*                                &  
&                                (du2(2:nx-1,2:ny-1,ARRAY_VAR3)      &  
&                                -du2(2:nx-1,2:ny-1,ARRAY_VAR3-1)) )
```

Fortran 90/95 Codes (cont.)

- Control of the array syntax conversion
 - done automatically
 - user can overwrite:
select an index dimension for conversion

